DOCUMENT RESUME

ED 092 576 TM 003 687

AUTHOR Wise, Helen
TITLE [Statement by Dr. Helen Wise, President, NEA, to the Symposium-"Statewide Educational Assessment: Coexistence or Confrontation"].
PUB DATE [74]
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974). For related documents, see TM003688 and 689

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS *Educational Accountability; *Educational Assessment; *Evaluation Criteria; *State Programs; Student Evaluation; Teacher Responsibility; Testing Problems; *Testing Programs
IDENTIFIERS *National Education Association

ABSTRACT

Although emphatically not against the concept of educational accountability, the National Education Association (NEA) feels that a redirection is needed in the implementation of such a system. Because of error, especially in testing minority and poor children, accountability programs should never use test results as the major source of data but should rely on multiple indexes. When testing is used, the NEA emphasizes the diagnostic capabilities of tests and warns against comparing students, schools or teachers. The NEA believes that teachers should be given the freedom to exercise professional judgment, to set learning goals for individual students, to assess the achievement of these goals and to establish the instructional procedures for attaining the desired learning. To expand and reinforce these comments, two NEA papers are included with this document: "Criteria for Evaluating State Education Accountability Systems" and "Testimony Presented by the National Education Association to the Panel on Evaluation of the Michigan Assessment Program." (RC)

ED 092576

TM 003 687

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

# STATEWIDE EDUCATIONAL ASSESSMENT: COEXISTENCE OR CONFRONTATION

A Symposium - AERA Convention

Chicago, April 19, 1972

Statement By

Dr. Helen Wise, President

National Education Association

STATEWIDE EDUCATIONAL ASSESSMENT: COEXISTENCE OR CONFRONTATION

The National Education Association is pleased to have been asked to participate in the symposium on Statewide Educational Assessment: Coexistence or Confrontation at the 1974 AERA Convention. The Association does have some definite points of view which we would like to present in behalf of our 1.4 million members and, we believe, for the benefit of students throughout the United States.

First, let me state emphatically that the NEA is not against the concept of educational accountability. We believe:

1. That teachers are and should be held accountable for professional competence - using the best educational practice available to meet the needs of students.

2. That as Daniel Griffith, Dean of Education, New York University, has said: "There is no social institution in the country as accountable as the public schools."

3. That testing, which is a focal point of accountability, when properly used is a valuable tool of education. Most school systems make extensive use of tests - both teacher made and commercially prepared.

4. In the right of the public, including policy makers, to full information about the status of education including both its strengths and its weaknesses.

5. That there are many things about most schools that need to be improved.

Then what's all the fuss about? Why the "confrontation?" Let me briefly state some issues the NEA is concerned about in the statewide assessment arena.

1. Some of the state accountability laws have noble goals but set up either faulty, or in some cases injurious methods of achieving their purposes. For example, using massive blanket testing procedures which are expensive and time-consuming when sampling techniques would meet the state goals equally well.

2. Other state accountability laws are satisfactory and have been supported by teachers' associations, but again many of these have failed to be implemented along acceptable lines. They have violated the spirit and purpose of the laws. They have imposed statewide goals and objectives and tended to emphasize skills that are easily measured at the expense of important learnings less easily measured. They have over-emphasized behavioral objectives requiring attention to minute details rather than broader humanistic goals.

3. We at the NEA believe that testing is but one way of obtaining evaluative data and one often in error, especially when measuring minority students or children of the poor. We believe that accountability programs should be based on multiple indexes, and that in no case should test results be the major datum.

4. When goals are set at the state level and attempts are made throughout the state to measure student achievement of these goals, there is interference in the local control of education. Local goals, objectives and needs of individual students are subjugated to the state goals which may or may not be appropriate for every student in every school. Often, too, the state is collecting data already available in local school districts.

5. We at the NEA have been very disturbed by some of the uses made of the results of state assessment programs. We believe that the results should be used to improve education, not to compare students, schools or teachers. The results

should be used for policy formulation, identification and diagnosis of student needs, identification of needed areas of in-service education of teachers, and the like. We are against the public disclosure of individual students' scores which we believe to be a violation of the privacy of the student.

6. We do not believe that the industrial input-output model is appropriate for education which is a humanistic endeavor with a multitude of condition variables entering in and over many of which the teacher and the school have no control.

7. We do not believe that the purpose of an accountability program should be to save money at the expense of a sound educational program for all students. Neither do we believe that the distribution of funds should ever be based on the results of tests no matter what type of tests are used.

8. We believe that teachers are frequently being held accountable for things over which they have no control. We believe that teachers should be given the freedom to exercise professional judgment, to set learning goals for individual students, to assess the achievement of these goals and to establish the instructional procedures for attaining the desired learning. There are many conditions, however, which influence learning which are beyond the control of the teacher such as the number of students which the teacher must deal with and the nature and amount of teaching material available. Teachers also are only one segment of the educational community which must assume responsibility in any accountability plan.

Yes, teachers are willing to assume a fair share of the responsibility for achievement of educational results. And we most certainly want to be involved in the

development and implementation of accountability plans including state assessment programs. Teachers take with high seriousness a committment to carry out their professional assignments in the most responsible manner they know how, in light of the many varying conditions under which they must perform.

We do not intend to stand by, however, and be made the goats for the failures of education. A surgeon is accountable for using the best surgical methods available. He cannot be held accountable for the patient's recovery. A teacher is accountable for professional competence - the knowledge and use of good educational practice - not in how much a student learns.

The question then is not "Statewide Educational Assessment: Coexistence or Confrontation" but rather "Statewide Educational Assessment: Needed Redirection."

To expand and reinforce these comments, two NEA papers are attached: Criteria for Evaluating State Education Accountability Systems and Testimony Presented by the National Education Association to the Panel on Evaluation of the Michigan Assessment Program.

ED 092577
TM 003 688

# CRITERIA FOR EVALUATING STATE EDUCATION ACCOUNTABILITY SYSTEMS

National Education Association

An acceptable accountability system should respect the complexity of reality. While a good system conceivably could improve education, a simplistic scheme could deal it deleterious blows and damage the lives of millions of children and teachers. Since children have little defense against ill-conceived schemes, it is incumbent upon professionals to examine such systems seriously.

Education is a serious enterprise. Its essence lies in what happens between a child and his parents, his teachers, and his classmates. These relationships are delicate and susceptible to strong outside influences, and an accountability system must take care not to damage them. Above all, the system must be "livable" for those who are expected to abide by it.

In a pluralistic society an accountability system should promote diversity, not conformity. Opportunities for diversity must exist for the child, the parents, the teacher, the school, and the community. Each entity has a right to be itself. A monolithic system which imposes a single set of values strikes at the very heart of individualism and democratic processes. In short, an accountability system should be responsive to individual differences. Ideally, such a system should strive for personalization.

A final axiom is that an accountability system must be judged on how it will function, not on what it promises. The consequences are far too great

for millions of children to rely on vague promises and glib promotion of what the system may do in the future. If the system is simplistic and primitive in its current form, it should not be implemented unless this is done in a carefully controlled laboratory setting where it can be tested and improved, and decisions of consequence to childrens' lives and teachers' careers should not be based on it.

In addition to these basic principles, there are a number of specific criteria for evaluating state accountability systems.

## STATED PURPOSES AND SPECIFIED USE OF RESULTS

1. <u>The purposes for which the state accountability system is to function should be clear, concise, and understanable to both the profession and the public</u>.

Such statements as "for statewide planning" or "for state decision-making purposes" are too vague to justify the large allocation of resources in time, money, and effort that are being pumped into many state accountability systems.

Stated purposes should describe and provide examples of how education is expected to be improved as a result of implementing the system. Justification on the basis of its contribution to implementing a program planning - budgeting - evaluation system is highly questionable. There is little evidence to date of a positive contribution by PPBES to educational improvement.

2. The uses to which data will be put as they result from the accountability system should be clearly spelled out in concrete, constructive, and positive terms.

Complete and detailed plans for uses of data resulting from accountability programs should be built in from the beginning. Initiating a program as extensive as is represented in most state accountability systems is a matter of high seriousness and potentially pervasive in its consequences. It should therefore be thought through fully and spelled out clearly and completely before implementation. And the detailing of how the resulting data are to be used should be clearly related to the stated purposes.

## PARTICIPATION AND CONTROL

3. Local control must be retained within a state accountability system.

Besides a long tradition of community control of education in the United States, a strong philosophical and pragmatic case can be made that most educational decisions are best handled close to those who must live with the consequences of those decisions. In education it is the teacher, the parent, and the child himself who are most appreciative of the complexities of the child's world and who are most vitally interested in his personal welfare.

The concrete and personal kinds of information required by the teacher and parent are quite different from the highly abstract and impersonal aggregated data demanded by state bureaucracies. Few would deny that the state

has a right to collect data to guide planning, but not at great expense to those at the local level. It would be a grim irony indeed if children's needs and local purposes were to suffer in order to serve the convenience of state administrators.

Just as states have found it necessary to defend many of their rights against the authority of the federal government, so local authorities are justified in defending against state authority their right to be treated as individual entities. Overall, an accountability scheme should not increase the centralization of governmental power.

4. Students, parents, and professionals who will bear the consequences of the accountability system should participate in its development and governance.

Participation means more than being consulted or serving on a committee. It means more than being caught up in an arrangement over which one has no influence. Participation means having influence on decisions and individual recourse to other action when those decisions are disagreeable. It may also ensure greater understanding and commitment.

5. The state accountability system should include explicit provision for holding the state departments of education in general--and state administrators in particular--accountable to local authorities and professionals.

If accountability can improve the local schools, it can also improve the state departments of education, which are often deficient in the eyes of many educators. Existing state legal responsibility and methods of

operation are not sufficient guarantees of accountability any more than is the legal responsibility of the local school board or the tenure status of teachers. A plan should be developed by which the public, students, and professionals can evaluate the competency of the state dep nt in carrying out its charges, preferably in the same manner as local schools and teachers are evaluated.

## DATA COLLECTION

The data collected on the effectiveness of the school must reflect the complexities of the educative process. This would eliminate accountability systems that pursue a few simplistic goals. An English statesman has said:

> As all policy makers know from experience, policy does not consist in prescribing one goal or even one series of goals; but in regulating a system over time in such a way as to optimize the realization of many conflicting relations without wrecking the system in the process. Thus the dominance of technology has infected policy-making with three bogus implications, just admissible in the workshop but lethal in the council chamber. One of these is the habit of accepting goals--states to be attained once for all--rather than norms to be held through time, as the typical object of policy. The second is the further reduction of multiple objectives to a single goal, yielding a single criterion of success. The third is the acceptance of effectiveness as the sole criterion by which to choose between alternative operations which can be regarded as means to one desired end. The combined effect of these three has been to dehumanize and distort beyond measure the high human function of the government--that is, regulation--at all levels.[1]

---

[1]Vickers, Geoffrey. Freedom in a Rocking Boat. Baltimore: Penguin Books, 1970.

6. The accountability system should provide for the collection of multiple-outcome data.

No educational program should be evaluated on the basis of a few pieces of information or one or two measures. Assuming that the complex purposes of the whole educational enterprise can be reduced to a few goals, such as teaching "basic skills," or evaluated on a single criterion, no matter what, has catastrophic portent. The more different kinds of data collected, the more likely the evaluation will reflect reality and be fair.

7. The system should provide data for assessing whether program elements and conditions are of a standard of quality to make possible high levels of performance by staff.

Up-to-dateness of curriculum, adequate materials and media, time to plan and to teach, reasonable teaching loads, availability of specialist and clerical services, opportunities for in-service education and decision-making power for teachers -- these and a number of other program conditions and arrangements effect the ability of school staffs to be accountable and must be taken into consideration.

8. The system should provide substantial information on what is going on in the classroom.

The classroom atmosphere in which a child spends a significant portion of his life is important, whether or not it results in increased learning. There have been documented instances recently in which children have been beaten to increase achievement scores. This practice is deplorable, even if it is effective. The humanness of the classroom should always be a

consideration of high priority.

9. The data collected should include professional judgments.

In the long run there are no substitutes for the reasoned and intuitive judgments of skilled and experienced educators. Social-system measures are too simplistic and primitive to be relied on exclusively. For example, in spite of great publicity, there are only a handful of studies on the efficiency of behavioral objectives, and these studies are equivocal. Ultimately we must rely on the human mind to make judgments. There are no mechanical substitutes.

10. The system should collect data by a variety of techniques from relevant groups and individuals.

Testimonials, interviews, classroom interaction analysis, opinion polls--all these provide a picture of the richness of educational life and mitigate against making decisions on inadequate and highly abstract information. It is especially important to find out what students are thinking and feeling; they are in the learning setting all the time and their observations are as reliable as any. Parents should also provide judgments on the quality of instruction.

## TESTS

No element of the accountability constellation is more inadequately understood by lay people than standardized achievement tests. The most ill-conceived of all accountability systems would be one which relies heavily on test

data gathered at the state level and reported publicly, and which rewards and punishes teachers and administrators on the basis of test results. The potentially destructive effects of such a system are mind-boggling.

11. Under no circumstances should standardized achievement test results be used as the major data in an accountability system.

Tests are not adequate and valid measures of what is taught in school. They are not responsive to school learning unless the teacher teaches the items on the tests. Since tests always sample a domain of behavior, teaching the items on a reading test, for example, does not necessarily mean that students have been taught to read.

The errors involved in testing have been thoroughly explicated by Stake.[2] While some test scores can be useful in diagnosing learning problems and assessing a child's progress, the practitioner also must use many other kinds of information in making decisions, (for example, student interviews, real and simulated performances, products of learning, student self-evaluation, student peer evaluation). When tests are used as a major criterion of learning, their deficiencies are glaring. For the most part they measure only recall-type tasks and shift teaching emphasis from complex mental abilities to those that are simple and easy to measure, this at the expense of long-term retention, relearning ability, and other learning considered by psychologists to be more important.

---

[2]Stake, Robert E. "Measuring What Learners Learn." School Evaluation: The Politics and Process. (Edited by Ernest R. House). Berkeley, Calif: McCutchan Publishing Corp., 1973. pp. 193-223.

12. If criterion-referenced tests are substituted for norm-referenced tests in order to overcome some of the testing problems, these tests should be closely scrutinized.

Since no one has been entirely successful in developing criterion-referenced tests, claims for their validity and reliability must be viewed with caution. One major criticism of norm-referencing that criterion-referencing has promised to correct is the lining up of students as related to a measure of central tendency (mean or median) which assures that half will be "below average" no matter how proficient they become in achieving instructional objectives. If movement is to be away from measures of central tendency, these new approaches to validity and reliability will be required.

In addition, averages and means should not be replaced by minimum competency levels, cutting scores, or pass-fail points. This will result only in replacing one statistical device with another for denying opportunities to some students and assuring them to others. It will frustrate the advantage attributed to criterion-referencing of being able to move all students, as rapidly as possible, toward full mastery without the deleterious effects of comparision with others.

13. Test results collected at the state level should not be publicized by school or school district.

Regardless of promises to the contrary, test data collected at the state level are almost invariably made public. The pressures are too great once the data are known to exist. Unless collected anonymously,

test results by school and school district should be known only at the local level. In any case, the plan for using the data should be clear in advance.

14. Reports should be tailored for different audiences in order to provide the highest understandability, to avoid misinterpretation, and to assure privacy.

When test scores are reported, the error of measurement should be communicated in understandable form, and only information that can be reported without infringing on the rights of individuals should be included. Information on an individual pupil should not be reported to anyone without the parents' consent. This is consistent with the principle that the district should be accountable to its own constituency rather than to the state agency.

15. If the state desires test data for its own planning purposes, it should use proven matrix sampling techniques which will not reveal schools and which will greatly reduce costs.

Matrix sampling techniques can give an accurate picture of the state by various categories much more efficiently than testing each child with an entire instrument. Otherwise, steps should be taken to protect individual identities. Carefully drawn samples are sufficient for state decision-makers.

16. Districts and schools should not be compared to one another on test scores, nor should a school be judged on the basis of achievement increments or decrements.

Achievement scores are highly influenced by the social and economic conditions within the school district or building attendance area. Therefore, comparisons have little meaning other than to indicate population characteristics. Calculating the gains between two administrations of a test is also highly dubious because of the large and irremediable errors of measurement, turnover in student population, and the like. If such gains are calculated at all, this should be done, not for individual students, but only for large groups in which errors can balance each other out.

17. Rewards or punishments should not be given on the basis of test scores, either group means or individual.

Test scores are highly subject to manipulation by teaching the tests, by selecting the time of year the tests are given, and by controlling conditions and instructions under which they are given. Moreover, tests were never constructed to provide an exact measure of where individuals stand at a given time. Test scores vary so much (called the "error of measurement") that it is possible for individual students to show great gains when no learning or instruction has occurred. The best tests may be a full-year grade equivalent in error. Under a system of rewards and punishments, the temptation to cheat is great indeed. In addition, the OED evaluation of

performance contracting has shown that test scores are not raised by rewarding on the basis of results. If a performance contractor is paid on the basis of individual student gains and not penalized for losses, he can make money simply on error fluctuations of test scores.

18. An accountability system should minimize dangerous side effects of relying on test results.

Side effects include suspicion, acrimony, cheating, and a wide assortment of potentially debilitating conditions. An accountability system that leads to wholesale cheating is likely to be counterproductive and result in poorer schools. And, when school administrators and teachers are put under pressure to produce specified results which cannot be assured, there is temptation to teach to the tests or teach the tests. Not only does such pressure mitigate against best professional performance, it encourages unethical acts.

19. If tests are deemed desirable at the local level, the local district should be able to choose among a set of commercially available tests or to develop their own criterion-referenced tests.

Test scores can be useful at the local level, by local option, without comparison with other districts other than what a program supplies. Tests selected or developed to reflect local goals and objectives can serve far more useful purposes than those that attempt to respond broadly to the student population of an entire state.

20. As a general rule, state agencies should not develop their own tests.

State education departments seldom have the manpower or competency to develop their own tests. Although the results of such efforts have not been promising, it would probably be more efficient to use those already available or, if absolutely necessary, to have tests developed on contract.

## COSTS IN DOLLARS, TIME AND PERSONNEL

21. The true cost of the accountability system should be calculated.

Often only a small part of the research and development money necessary to initiate the system is included in costs. The true costs must include the time professionals, children, and others spend providing data. It has been estimated that a complete testing program for a large state, if properly developed and implemented, would cost tens of millions of dollars.

22. The accountability system should not overload professionals or children with providing data.

Many federal programs require the work of a number of local staff to fill out forms, as well as those at the state end of the system to record and analyze the data provided. An accountability system should not be burdonsome to administrators, teachers, or children.

23. The accountability system should require a minimum number of people in the state bureaucracy.

Many state education agency personnel are building careers on the accountability movement. The larger the state bureaucracy becomes in this area, the more it will serve as a lobby to expand continually its own operations past the point of useful returns. The result will be empire building for its own sake.

24. An accountability system should have explicit provisions for the evaluation of its processes and effects.

An accountability program should itself be accountable through a comprehensive plan for auditing its processes, results, and their usefulness for educational improvement.

25. Plans for auditing[3] the success of accountability programs in accomplishing their stated purposes, should be built in as the program is planned.

Auditing programs that are tacked on as afterthoughts or are developed after the program is under way are likely to be ineffective. They appear to suffer inefficiencies, analogous to those in auto air-conditioners installed after the car has left the factory -- they don't respond directly to the nature of the operating mechanism.

---

[3]The term auditing is used here in the context of evaluating the evaluation.

26. <u>There should be several kinds of audits applied to a state account-ability system</u>.

Certainly the state itself should plan early for its own intensive evaluation of its accountability efforts by a variety of criteria. Foremost among these should be the criterion of how and how much the program has contributed to improving education of the children in the state. In addition, independent outside audits are mandatory. The first line of consideration for such audits should be the teachers of the state. They are the ones who will in the final analysis implement the program and whose professional performance will be influenced most by it. And it is their expertise and professional judgment that should count most. Finally, an outside agency competent in applying the most sophisticated evaluation tools, which is the most independent and impartial and that has the highest credibility to both professionals and the public. should be retained. Such multiple-index and multiple-agency evaluations should take place no less than annually.